JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 1993, 60, 495-514 NUMBER 3 (NOVEMBER)

SIGNAL-DETECTION PROPERTIES OF VERBAL SELF-REPORTS

THOMAS S. CRITCHFIELD

AUBURN UNIVERSITY

The bias (B'_{H}) and discriminability (A') of college students' self-reports about choices made in a delayed identity matching-to-sample task were studied as a function of characteristics of the response about which they reported. Each matching-to-sample trial consisted of two, three, or four simultaneously presented sample stimuli, a 1-s retention interval, and two, three, or four comparison stimuli. One sample stimulus was always reproduced among the comparisons, and choice of the matching comparison in less than 800 ms produced points worth chances in a drawing for money. After each choice, subjects pressed either a "yes" or a "no" button to answer a computer-generated query about whether the choice met the point contingency. The number of sample and comparison stimuli was manipulated across experimental conditions. Rates of successful matching-to-sample choices were negatively correlated with the number of matching-to-sample stimuli, regardless of whether samples or comparisons were manipulated. As in previous studies, subjects exhibited a pronounced bias for reporting successful responses. Self-report bias tended to become less pronounced as matching-to-sample success became less frequent, an outcome consistent with signal-frequency effects in psychophysical research. The bias was also resistant to change, suggesting influences other than signal frequency that remain to be identified. Self-report discriminability tended to decrease with the number of sample stimuli and increase with the number of comparison stimuli, an effect not attributable to differential effects of the two manipulations on matching-to-sample performance. Overall, bias and discriminability indices revealed effects that were not evident in self-report accuracy scores. The results indicate that analyses based on signal-detection theory can improve the description of correspondence between self-reports and their referents and thus contribute to the identification of environmental sources of control over verbal self-reports.

Key words: self-reports, matching to sample, signal detection, discriminability, bias, signal-frequency effects, button press, button release, college students

---

Language and communication are frequently studied empirically (e.g., R. Brown, 1970; Klima & Bellugi, 1979; Lennenberg, 1967; Vygotsky, 1978; Walker & Blaine, 1991) but rarely in the context of the experimental analysis of behavior (McPherson, Bonem, Green, & Osborne, 1984; Oah & Dickinson, 1989). The verbal self-report exemplifies this state of affairs. As a common form of instrumentation, self-reports are of exceptional interest to clinical psychologists, cognitive psychologists, and researchers of behavior that tends not to occur publicly (e.g., sexual practices or illicit drug use). Quite often, however, the referent events of interest to many researchers make corroboration of the self-reports—presumably a necessary component of an experimental analysis—problematic.

Some have approached the problem of corroboration by proposing phenomenon-specific theoretical principles to guide the interpretation of uncorroborated self-reports (e.g., Ericsson & Simon, 1984). Because these principles both derive from and explain uncorroborated self-reports (e.g., reports about private events), however, it is unclear how their validity should be evaluated (e.g., Hayes, 1986). An alternative approach is to view the verbal self-report—a response presumably under discriminative control of characteristics or actions of the person making the report—as behavior subject to the same fundamental influences as any other. This approach encourages the study of self-reports in behavioral assays created for scientific advantage (including easy corroboration), with a long-range goal of generalizing to situations in which corroboration is not possible (Critchfield & Perone, 1993). Systematic research of this type has not been especially common, but reason for optimism can be found in procedures that have been developed in several different research traditions (e.g., Kausler & Phillips, 1988; Shimp, 1981).

I thank Larry Alferink and Peter Harzem for use of laboratory space, Michael Perone for use of equipment, Shirley Springer for data collection assistance, and Scott D. Lane and M. Christopher Newland for comments on the manuscript. Portions of the study were described in 1992 at the 18th Association for Behavior Analysis Convention in San Francisco, and at the 25th International Congress of Psychology in Brussels, Belgium. Lea T. Adams coauthored the latter presentation. Address reprint requests and correspondence to the author at the Department of Psychology, Auburn University, Auburn, Alabama 36849.

		DMTS RESPONSE	
		Successful	Unsuccessful
SELF-REPORT	"I succeeded"	HIT	FALSE ALARM
	"I failed"	MISS	CORRECT REJECTION

Fig. 1. Contingency matrix showing self-reports of DMTS success as a function of actual success; cells are labeled using response classes derived from signal-detection theory.

Once methods are devised to allow the objective corroboration of self-reports, the practical issue arises of how best to describe correspondence between the reports and their presumed referents. Accuracy scores, although often employed for this purpose (e.g., Critchfield & Perone, 1990b; R. Nelson, 1977; Shimp, 1981), provide an overly broad picture of correspondence that masks much information, including whether inaccurate self-reports reflect a failure to report the occurrence or nonoccurrence of the referent event.

In the analysis of self-reports about behavior, a more precise strategy is to employ a two-by-two matrix of occurrences and nonoccurrences familiar in signal-detection theory. In some signal-detection procedures, a signal (such as a tone or a light) occurs on some trials and not on others; on each trial the subject reports its presence or absence. The conjunction of these events creates four possible response categories defined in terms of the status of the signal and the "content" of the report. Simple self-reports can be analyzed within this framework if the "signal" is a response made by the reporter rather than an external stimulus.

Figure 1 illustrates the approach used in the present experiment. The behavioral "signal" is a response that meets the reinforcement contingency of a delayed matching-to-sample (DMTS) procedure; a successful referent response may occur or not occur on a given trial. Similarly, a self-report describing the events of each trial may indicate that a successful response either did or did not occur. Each of the four resulting combinations represents a different relationship between referent and self-report and may be labeled using terms coined to describe analogous relations in the reporting of external events (e.g., Green & Swets, 1966). Moreover, rates of the four self-report categories can be used to calculate formal indices of response bias and discriminability (e.g., Grier, 1971).

In a study manipulating the number of distractor items in a DMTS sample-stimulus display, Critchfield and Perone (1993) found that accuracy scores inadequately described patterns of self-reports about DMTS success. The analytical strategy shown in Figure 1 revealed additional effects, including a positive correlation between the number of DMTS sample stimuli on each trial and self-report discriminability (A'; Grier, 1971), and a preponderance among inaccurate self-reports of false alarms (inaccurate reports of success) over misses (inaccurate reports of failure). The latter tendency was described quantitatively in terms of a pronounced bias (B'_H; Grier, 1971) for reporting success and proved to be pervasive, occurring in 89 of 90 experimental conditions across 6 subjects.

The consistent "report-success" bias observed by Critchfield and Perone (1993) may be attributable in part to the fact that the referent event, DMTS success, typically occurred on more than 50% of the trials. Signal frequency is a common source of response bias in signal-detection paradigms (Gescheider, 1985). For example, in a test situation involving the presence or absence on each trial of a weak tone, bias scores tend to reflect the relative frequency of presences versus absences. With the tone present on a large majority of trials, bias scores typically indicate a predisposition for reporting the presence, rather than the absence, of the signal. Analogously, subjects in the Critchfield and Perone study may have been predisposed to report DMTS success because that event occurred so frequently. If so, the report-success bias could be considered to be an artifact of the test situation and would not appear at lower rates of DMTS success. From this perspective, it is interesting that subjects in the Critchfield and Perone study showed a weak tendency for the report-success bias to become less pronounced as DMTS success became less frequent, as might be anticipated from signal-frequency effects.

The purpose of the present investigation was to extend Critchfield and Perone's (1993) preliminary characterization of bias and discriminability in self-reports about DMTS success. The possible situational nature of a report-success bias was examined by engineering DMTS success rates lower than 50%. As in several previous studies, the referent response occurred in a DMTS task in which points were contingent on selection of the matching comparison stimulus within a time limit (Critchfield & Perone, 1990a, 1990b, 1993). After each DMTS trial, subjects self-reported by pressing "yes" and "no" buttons to answer a computer-presented query about the success of the last response in meeting the point contingency.

Critchfield and Perone (1993) manipulated DMTS success by varying the number of stimuli in a DMTS sample compound from two to four, while holding constant the number of comparison stimuli at two. In the present study, DMTS success was manipulated both via the number of sample stimuli (one to four) and via the number of comparison stimuli (two to four). Experimental conditions were defined by the number of sample and comparison stimuli presented on each trial. For example, in a condition with three samples and two comparisons, only one of three samples would recur among the two comparisons. The other two comparisons, and one sample, would be irrelevant or distractor items.

This means of manipulating DMTS success was convenient because it permitted a comparison of effects related to manipulating the number of sample stimuli with effects related to manipulating the number of comparison stimuli. Thus, the design permitted self-report patterns to be viewed in two ways: as a function of overall DMTS success rates and as a function of the number of nonmatching sample and comparison stimuli. This flexibility proved to be valuable, because the results showed self-report bias to be better characterized as a function of DMTS success rates and self-report discriminability to be better characterized as a function of the number and location in the DMTS trial of nonmatching stimuli.

METHOD

Subjects

Two male (S2 and S4) and 8 female undergraduate students volunteered to participate in a laboratory experiment on "human performance and decision making." Subjects received bonus credit in psychology classes based on their hours of participation; during sessions, they accumulated points that served as chances in a drawing for cash prizes.

Apparatus

Subjects worked alone in a small room containing a table, chair, and a response console with a monochrome video monitor resting on it (for details, see Critchfield & Perone, 1990b). Subjects performed the DMTS task using four round illuminable response keys arranged horizontally near the bottom of the console's sloping front panel. Self-reports were made using two push buttons, each mounted to a small box extending from one side of the console. A microcomputer outside the workroom controlled experimental events and collected the data.

Procedure

Trial format. The procedure was based closely on that of Critchfield and Perone (1993). During the main experiment, each trial consisted of one DMTS response followed immediately, when scheduled, by a self-report, feedback about the success of the DMTS response, and consequences contingent on the self-report. Trials were separated by an intertrial interval (ITI) lasting at least 1 s. Subjects initiated each trial at the end of the ITI. This ensured that a subject was oriented toward the video screen when stimuli were presented, but also meant that the ITI could extend beyond its nominal value.

The video screen was divided into an upper box, used in conjunction with the DMTS task, and a lower box, used in conjunction with the self-report portion of the trial. At the start of each trial, the four buttons on the front of the console became illuminated and the message "HOLD LIGHTED BUTTONS DOWN" appeared in the center of the upper box on screen. Simultaneously depressing all four buttons cleared the message and produced, in the center of the DMTS box on screen, a sample-stimulus display lasting 800 ms. Subjects typically used the thumb and index finger of each hand to depress the buttons, which remained depressed until used to select a comparison stimulus. Panel A of Figure 2 shows one possible sample-stimulus display (construction of

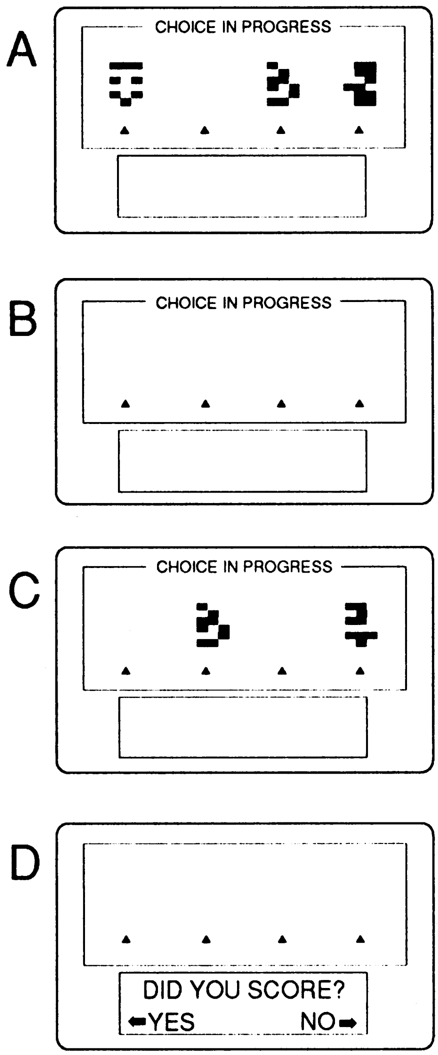

Fig. 2. Summary of the subject's display. Panels A, B, and C illustrate events during the DMTS trial, including sample-stimulus presentation, the intertrial interval, and comparison-stimulus presentation. Panel D shows the prompt used to generate self-reports.

the stimuli is described below). Following a 1-s delay (Panel B), comparison stimuli appeared in locations corresponding to at least two of the depressed buttons (Panel C). One comparison stimulus matched one sample element, and the others were randomly generated. Subjects attempted to select the matching comparison stimulus by releasing the round button corresponding to it. A successful response was recorded if a correct choice occurred within a time limit, normally 800 ms after presentation of the comparison stimuli. No stimulus change indicated when the time limit had elapsed.

Immediately after the choice, the DMTS box on the screen cleared and the center of the self-report box displayed the query, "Did you score?" (the word "score" had been used during preliminary training to signal point delivery). Below it, the labels "←YES" and "NO→" appeared 1 cm from the right and left sides of the self-report box, respectively (Panel D of Figure 2). Pressing the button attached to the console's left side registered a "yes" report, and pressing the button attached to the console's right side registered a "no" report. Pressing either of these side buttons cleared the screen and advanced the trial to the next scheduled event.

When scheduled, feedback about the success of the DMTS response immediately followed the self-report. Three feedback messages appeared simultaneously for 1 s in the DMTS area of the screen. The first message stated, "Your choice was CORRECT [or WRONG]." The second message stated, "Your choice was FAST ENOUGH [or TOO SLOW]." The third message summarized the implications of the other messages for point reinforcement, stating either "YOU SCORED! x points added to your total," or "NO SCORE" (x = 1 or 2, depending on the session; see below). When no DMTS feedback was scheduled, the trial advanced immediately to the next event.

When scheduled, a 1-s message describing the accuracy and point consequences of the self-report occurred next. In the self-report area of the screen a message stated, "RESULTS OF YOUR REPORT," accompanied by either "Correct—x point bonus" or "Wrong—x point penalty" as appropriate to the preceding self-report (x = either 1 or 3 points, depending on the session; see below under Session and Condition Format). When no feedback about self-reports was scheduled, the trial advanced immediately to the ITI.

Throughout the trial, error messages discouraged responses not conforming to the experimental protocol. For example, release of DMTS buttons before comparison stimuli were presented produced a message stating "Illegal Action!" and caused the trial to begin again with new stimuli. If a non-self-report button was depressed during the self-report query, a 2-s message stated "Illegal Action!" and the self-report query was presented again. For further details of error messages, see Critchfield and Perone (1990b).

DMTS stimuli. Figure 2 shows examples of the stimuli. Each sample and comparison stimulus consisted of a six-by-three matrix of rectangular cells, of which as few as three or as many as 18 could be illuminated (similar stimuli were described by Baron & Menich, 1985). An element could be as large as 10 mm by 7 mm, depending on how many cells were illuminated. On each trial, stimuli were drawn randomly from a pool of several thousand unique shapes, without replacement except for the obvious exception that one sample stimulus always matched one comparison stimulus.

Across conditions, the number of sample stimuli displayed on each trial ranged from one to four, and the number of comparison stimuli from which subjects chose ranged from two to four. For example, in one condition, three sample stimuli were presented, one of which subsequently appeared among two comparison stimuli (Figure 2). In another condition, a single sample stimulus was presented and subsequently appeared among four comparison stimuli. As shown in Figure 2, each sample and comparison stimulus was displayed in one of four possible locations within the DMTS box on screen. Stimulus locations were approximately 2 cm apart, each underscored with a small illuminated dot. The entire stimulus array appeared centered within the DMTS box.

If the number of sample or comparison stimuli was less than four, unused locations were left blank (e.g., Panels A and C of Figure 2). On such occasions, the locations actually used were randomly determined on each trial. During comparison display, each of the four locations corresponded to one of the round illuminated buttons being depressed on the console. Subjects indicated their choice of a comparison stimulus by releasing the button corresponding to the location of the matching stimulus. If the button released corresponded to an unused stimulus location, the trial was canceled, the screen cleared, and a 4-s message stated, "Illegal action! You cannot choose a blank." The trial then restarted using new stimuli.

Session and condition format. In sessions lasting 100 trials (about 8 to 12 min), 50-trial blocks were separated by a 20-s intermission, during which the screen was blank except for a message stating, "Intermission—Please wait." Subjects usually completed eight sessions during each 2-hr visit to the laboratory, allowing for brief subject-initiated rest periods between the sessions. At the end of each session, a message on the subject's screen displayed the number of points accumulated during that session. The message included an overall session total and subtotals reflecting the number of points (out of 100) earned from DMTS and the number of points accumulated from self-reports. The self-report total was further broken down into total point gains and total point losses.

Each experimental condition lasted eight sessions. Sessions 1 through 3 consisted solely of DMTS trials without self-reports; each DMTS response was followed by the outcome feedback described previously. Successful DMTS responses (those that were both correct and faster than the time limit) earned 2 points. Session 4 was intended to enhance the correspondence between self-reports and DMTS outcomes. Each DMTS choice was followed by a self-report and then feedback messages describing, in sequence, the success of the DMTS response and the consequences of the self-report. Successful DMTS responses earned 1 point. Accurate self-reports earned 1 point, and inaccurate ones resulted in a 1-point deduction from the subject's total. Sessions 5 through 8 provided the main data for the experiment and differed from the fourth session in only two respects. No feedback messages described DMTS performance after any trial, and self-report consequences operated on a random-ratio (RR) 3 schedule. To hold relatively constant the number of points that potentially could be earned in a session, an accurate self-report produced a gain of 3 points and an inaccurate self-report produced a loss of 3 points.

Instructions. Subjects read the following printed instructions just before the first session (ellipses indicate that nonessential information or elaboration has been omitted for brevity).

> In front of you is a console containing several lights and buttons. Your job is to make decisions based on information presented on your screen, and to indicate your decisions using buttons on the console. . . . When you depress the lighted round buttons, one or more "sample" shapes will appear briefly for you to study, then dis-

Table 1

Summary of experimental conditions. Conditions were defined, and named, according to the number of sample and comparison stimuli presented on each delayed matching-to-sample trial. In condition names, the first digit indicates the number of samples, and the second digit indicates the number of comparisons. Conditions experienced by all subjects are shown in boldface.

Number of comparison stimuli	Number of sample stimuli			
	1	2	3	4
2		**22**	**32**	**42**
3	13	**23**	**33**	**43**
4	14	**24**	**34**	44

appear. Shortly afterward, some "test" shapes will appear. Your job is to decide which one of these test shapes matches one of the samples. You can indicate your decision by releasing the lighted button corresponding to the matching shape (note that there are four positions on your screen, and four lighted buttons). . . . Note that you must hold down the lighted buttons until you are ready to indicate your decision. If you release too soon, the trial will cancel and start again, wasting time in which you could be earning points. You can earn points each time you choose the correct (matching) test shape. In order to earn a point, your choice must be both correct and within a time limit. . . . During your first session, you will have a relatively long amount of time to make each decision. Thereafter, the time limit will become more stringent. Do the best you can under the time constraints. The time limit will not change after your first work day. To begin with, after each complete trial, messages on your screen will tell you whether you earned points. Later on, you may be given less or different information about your decisions. Your screen will give you new instructions if the way you earn points should change. . . . Do not attempt to ask questions or leave the room until the work period is over. . . . Beyond the information contained in these instructions, it is up to you to decide how to operate the console to your best advantage. This is all the information we can provide at this time. If you have any questions, please ask them now.

Each session began with messages on the computer screen describing the point contingencies operating in that session. For "matching decisions," the message stated the number of points earned per "score" and whether scores would be signaled on screen. For "reports," the message stated the point value of each "bonus" and "penalty," and noted that point consequences for self-reports occurred only when indicated by feedback messages on the screen. Subjects cleared these messages and began the session by pressing a button located near the top of the console's front panel.

Preliminary training. An eight-session preliminary training phase using the format just described was designed to familiarize subjects with the DMTS task and the self-report procedure. Preliminary training differed from normal experimental conditions in three ways. First, the stimulus pool consisted of 13 keyboard characters (e.g., #, >, and &). Second, feedback messages lasted 2 s instead of 1 s. Third, at the beginning of the first session, the time limit for DMTS choices was 3,000 ms, and decreased across blocks of 50 trials according to the following sequence: 2,000, 1,000, and 800 ms. Thus, by the middle of the second session, the time limit had reached its typical value for the experiment (800 ms). DMTS trials always consisted of two sample stimuli and three comparison stimuli.

Experimental conditions. Conditions were defined, and named, according to the number of sample and comparison stimuli appearing on each trial. For example, when three sample stimuli were presented, one of which appeared among two comparison stimuli (as in Figure 2), the condition was designated as "32," with the first digit describing the number of samples and the second the number of comparisons. Table 1 delineates the stimulus configurations used in each condition. Each subject participated in a different sequence of at least eight experimental conditions (Table 2) selected to produce a broad range of DMTS success rates. As noted previously, the time limit on DMTS responding was 800 ms, but on two occasions (Condition 44 for S4 and S8), the normal time limit of 800 ms was reduced (to 700 and 450 ms, respectively), in an attempt to produce low DMTS success rates.

Previous research suggested that performance (both DMTS and self-reports) would stabilize within the number of trials allotted to each experimental condition (e.g., Critchfield & Perone, 1993), an assumption that generally was borne out. Overall percentages of successful DMTS responses and accurate self-reports from the final four sessions per condition were divided into blocks of 50 consecutive trials and subjected to a post hoc stability test in which the difference between mean per-

Table 2

Sequence of conditions for each subject. Condition names reflect the number of stimuli in delayed matching to sample, with the first digit indicating the number of sample stimuli and the second digit the number of comparison stimuli.

Subject	Order in sequence of conditions										
	1	2	3	4	5	6	7	8	9	10	11
S1	23	14	32	34	22	44	24	42	43	33	
S2	23	13	34	24	42[a]	22	32	43	33	42	
S3	23[a]	33	13	34	14	32	42	24	22	43	23
S4	23[a]	34	14	32	44[b]	24	22	42	33	43	23
S5	24	43	22	33	23	34	42	32			
S6	32	23	34	43	22	33	24	42			
S7	22	42	33	13	23	24	43	32	34		
S8	32	24	43	22	33	42	23	34	44[c]		
S9	23	33	42	32	43	34	22	24			
S10	23	22	42	34	32	24	33	43			

[a] Data lost due to computer malfunction; condition repeated at end of sequence.
[b] Time limit = 700 ms.
[c] Time limit = 450 ms.

centages in the first and second four blocks was considered as a proportion of the eight-block grand mean. For DMTS success, this proportion was less than .15 in 85% of the cases. More than half the cases in which the proportion was higher occurred in 3 subjects (S1, S9, and S10). For self-report accuracy, the proportion was less than .15 in 95% of the cases.

RESULTS

Data for each subject were summed across the final four sessions (400 trials) per condition prior to analysis. The results describe self-report patterns first as a function of the rate of successful DMTS referent responses in each condition and second as a function of the number of DMTS sample and comparison stimuli. In both cases, the data describe DMTS success (to show the behavioral context in which self-reports occurred), rates of self-report errors, the bias (B'_H) and discriminability (A') of the self-reports, using nonparametric indices from Grier (1971). A third set of analyses examines the relationship between self-reports and the response characteristics (speed and accuracy) that determined DMTS success.

Self-Reports As a Function of DMTS Success Rate

Experimental conditions were selected partly to produce a broader range of DMTS success rates than in previous self-report studies, including values lower than 50%. Table 3 shows, for each subject, the percentage of trials on which the DMTS response was successful, that is, on which the matching comparison stimulus was selected within the time limit. The top row of data for each subject shows that the manipulation of DMTS sample and comparison stimuli across conditions did produce a variety of success rates, from a median low of 33% to a median high of 88%. Success rates were lower than 50% in at least one condition for all subjects, although only marginally so for S5 and S6. The bottom two rows of data for each subject describe the specific response characteristics on which DMTS success was contingent—speed (percentage of responses faster than the time limit) and accuracy (percentage of correct DMTS responses). As in previous studies of DMTS performance under conjunctive speed-accuracy contingencies (Baron & Menich, 1985; Critchfield & Perone, 1990a, 1990b, 1993), variations in overall DMTS success reflected changes in both speed and accuracy.

Figure 3 shows the percentage of total self-report errors plotted as a function of the DMTS success rate in each condition. The figure is included primarily to show that an analysis based on overall self-report error rates (e.g., accuracy) can be uninformative. Although for some subjects self-report error rates were negatively correlated with DMTS success rates (e.g., S8 and S9), overall there was little systematic relation between the two variables.

Although rates of total self-report errors bore no systematic relation to DMTS success rates, the relative frequencies of two types of self-

Table 3

Percentage of delayed matching-to-sample responses that were correct, faster than the time limit, and both (successful). See Table 1 for key to experimental condition names.

Subject	Response characteristic	Experimental condition										
		13	14	22	23	24	32	33	34	42	43	44
S1	successful			89	57	66	67	57	38	60	47	37
	fast enough			95	63	80	91	79	63	88	74	71
	correct			93	88	80	73	69	55	67	60	53
S2	successful	88		83	46	57	72	60	37	67	42	
	fast enough	90		100	59	77	98	86	64	97	82	
	correct	94		83	72	69	73	72	52	69	48	
S3	successful	92	92	88	72	63	68	38	30	78	50	
	fast enough	94	96	92	79	70	78	44	34	89	73	
	correct	97	95	94	84	81	84	70	61	75	62	
S4	successful		87	91	80	64	76	67	36	72	53	26
	fast enough		96	98	91	81	73	92	46	97	95	32
	correct		91	93	86	81	94	72	70	74	58	61
S5	successful			88	73	49	77	59	50	75	41	
	fast enough			96	89	65	93	80	67	97	69	
	correct			92	81	70	82	70	68	77	59	
S6	successful			79	45	56	52	50	26	64	32	
	fast enough			86	64	62	73	77	48	94	55	
	correct			88	65	78	71	62	50	67	54	
S7	successful	92		72	64	52	71	48	49	42	45	
	fast enough	95		92	90	83	93	89	77	42	89	
	correct	96		79	71	60	76	53	57	49	52	
S8	successful			88	77	58	66	58	50	60	48	25
	fast enough			97	87	84	95	89	75	95	87	36
	correct			91	86	68	68	64	61	63	53	49
S9	successful			85	60	71	64	44	43	61	35	
	fast enough			90	70	85	78	64	57	84	53	
	correct			93	79	79	80	58	62	69	56	
S10	successful			67	41	37	52	45	20	34	38	
	fast enough			68	50	40	62	55	24	39	65	
	correct			83	80	78	63	67	55	68	54	

report errors did. Figure 4 shows rates of false alarms and misses as a function of the DMTS success rate. False-alarm rates reflect the number of false alarms in each condition divided by the sum of false alarms and correct rejections. Miss rates reflect the number of misses in each condition divided by the sum of misses and hits. In most conditions for most subjects, false-alarm rates were higher than miss rates. As DMTS success rates increased, false-alarm rates tended to increase and miss rates, already low, tended to decrease.

Figure 5 shows self-report discriminability (open circles plotted against the right ordinate) and self-report bias (filled circles plotted against the left ordinate) as a function of DMTS success rates. For the present data set, the A′ index of discriminability estimates an individual's detection of a "signal" consisting of a successful DMTS response. Values can range from 0 to 1.00, with .50 indicating chance levels of detection and increments between .50 and 1.0 indicating increased discriminability. Discriminability scores were not systematically related to the DMTS success rate.

In the context of the present study, the B'_H index of bias estimates an individual's relative tendency to report successful or unsuccessful DMTS responses, independent of the actual success of the DMTS response. Values can range from -1.0 to $+1.0$, with negative values representing a bias toward reporting success and positive values representing a bias toward reporting failure. A score of 0 indicates no bias. Bias scores were clustered near the negative end of the bias scale, indicating that under most

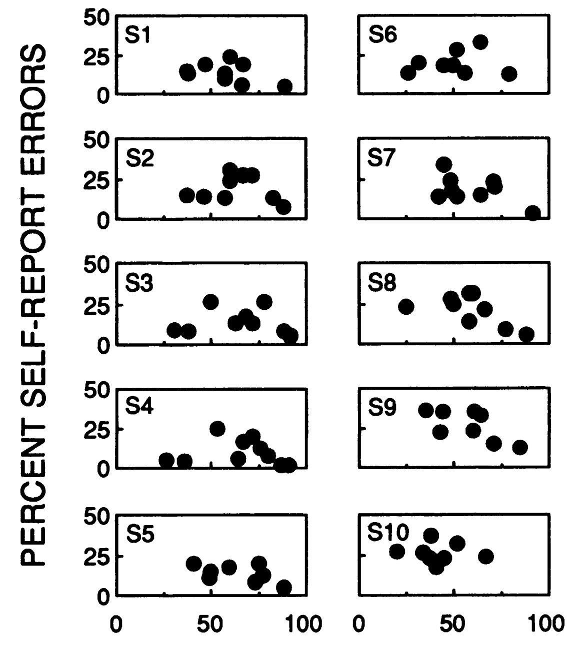

Fig. 3. Percentage of total self-report errors as a function of the percentage of successful DMTS responses.

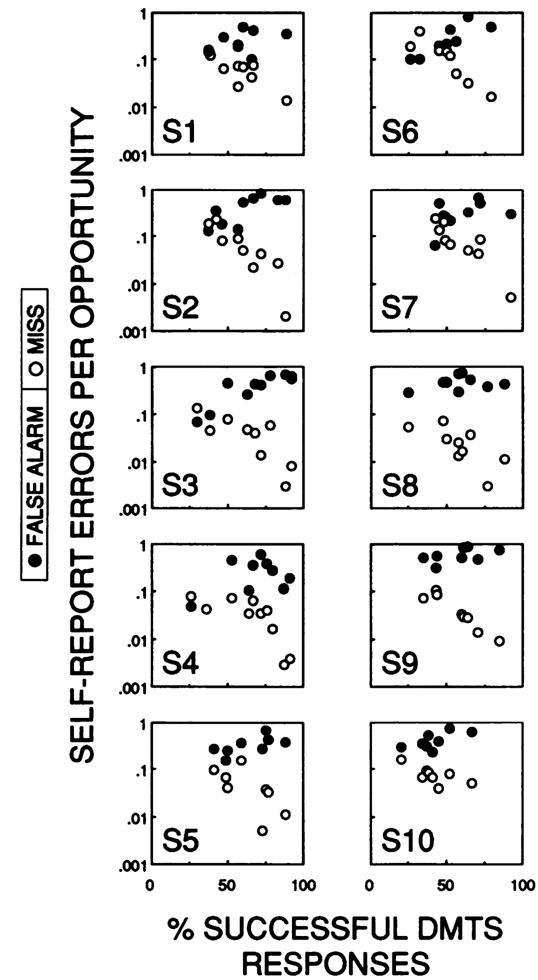

Fig. 4. Self-report errors: Misses and false alarms per opportunity as a function of the percentage of successful DMTS responses.

circumstances subjects exhibited a bias for reporting DMTS success. For all 10 subjects, however, bias tended to become less pronounced as DMTS success became less frequent. In 5 subjects (S2, S3, S4, S6, and S7) a bias for reporting failure occurred at low DMTS success rates. Thus, the bias functions appeared to show evidence of signal-frequency effects common in other types of signal-detection tasks.

Bias functions due solely to signal frequency would be expected to cross the zero-bias threshold at a point at which occurrence and nonoccurrence of the "signal" are equally probable—in this case at 50% DMTS success, a location marked with a small cross in the center of each panel in Figure 5. To facilitate comparisons between this hypothetical crossover point and actual bias functions, the bias data for each subject were fitted via either least squares linear (S5, S8, and S9) or logarithmic (the remaining subjects) regression, whichever accounted for the greater proportion of variation (these proportions ranged from a low of .52 for S7 to a high of .95 for S4, with a median of .67). Noteworthy is the fact that the individual-subject functions each pass to the left of the hypothetical crossover point or fail to cross the zero-bias threshold at all (the same holds true when logarithmic fits are used for S5, S8, and S9). Thus, biases for reporting failure, if they occurred at all, occurred only at DMTS success rates below 50%, suggesting that the report-success bias was resistant to change in a fashion not predicted solely by signal frequency.

Self-Reports As a Function of the Number of DMTS Sample and Comparison Stimuli

Table 1 (conditions in boldface) shows that this study may be thought of as a factorial

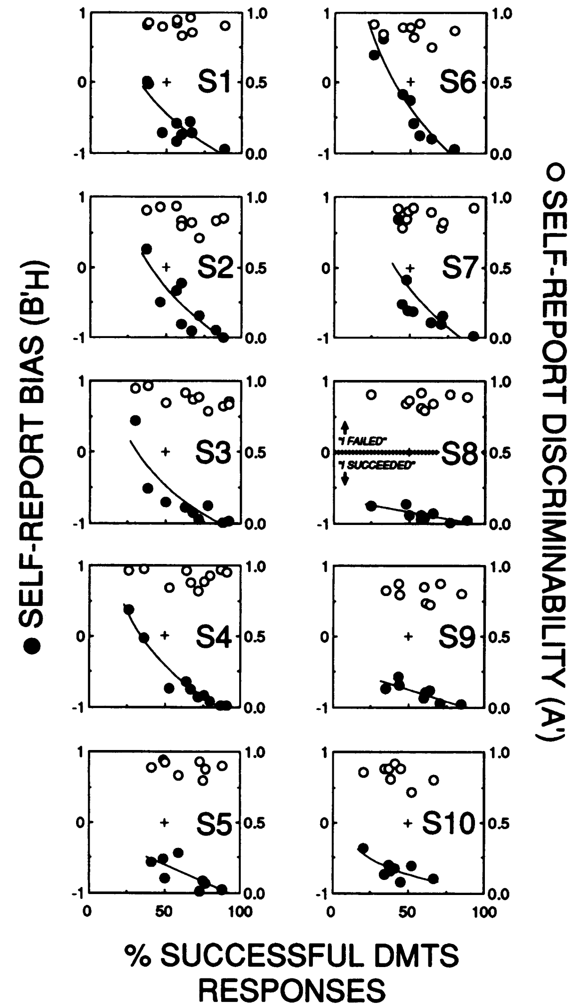

Fig. 5. Self-report bias (plotted on the left ordinate) and self-report discriminability (plotted on the right ordinate) as a function of the percentage of successful DMTS responses. For bias, positive scores indicate a predisposition for reporting failure; negative scores indicate a predisposition for reporting success.

design with number of DMTS sample stimuli and number of DMTS comparison stimuli as within-subject factors, although straightforward factorial analysis of the eight conditions in which all 10 subjects participated is precluded by a missing condition (44) in which not all subjects participated. As a consequence, different subsets of the data were examined to explore the effects on self-reports of the number of DMTS sample stimuli and of the number of DMTS comparison stimuli. In Figures 6 through 8, one set of six conditions (22, 32, 42, 23, 33, and 43) was used to compare the effects of three levels of sample-stimulus number (two, three, four); separate functions show these effects when there were two and three comparison stimuli. A different set of six conditions (22, 23, 24, 32, 33, and 34) was used to compare the effects of three levels of comparison-stimulus number (two, three, and four); separate functions show these effects when there were two and three sample stimuli. Because of substantial intersubject variability on some measures, repeated measures analyses of variance were used to corroborate the patterns shown graphically (Appendix).

The left portion of Figure 6 illustrates the analytical strategy, as applied to DMTS success rates. Each panel contains 10 functions, one for each subject. The two panels at the top left show DMTS success as a function of the number of sample stimuli (two, three, or four). Panel A includes conditions with two comparison stimuli (Conditions 22, 32, and 42), and Panel B includes conditions with three comparison stimuli (Conditions 23, 33, and 43). Thus, for statistical purposes, the top two panels constitute a three (number of samples) by two (number of comparisons) factorial design. DMTS success tended to decrease with the number of sample stimuli. The effects shown graphically were corroborated statistically as significant main effects for number of samples and number of comparisons; the samples-by-comparisons interaction was nonsignificant (Appendix).

The two panels at the bottom left of Figure 6 show DMTS success as a function of the number of DMTS comparison stimuli. Panel C includes conditions with two sample stimuli (Conditions 22, 23, and 24), and Panel D includes conditions with three sample stimuli (Conditions 32, 33, and 34). Thus, for statistical purposes, the bottom panels constitute a three (number of comparisons) by two (number of samples) factorial design. DMTS success tended to decrease with the number of comparison stimuli. The effects shown graphically were corroborated statistically as significant main effects for number of samples and number of comparisons; the samples-by-comparisons interaction was nonsignificant (Appendix).

Of particular interest is the comparison of Panels A and B (those showing the influence of sample-stimulus number) with Panels C

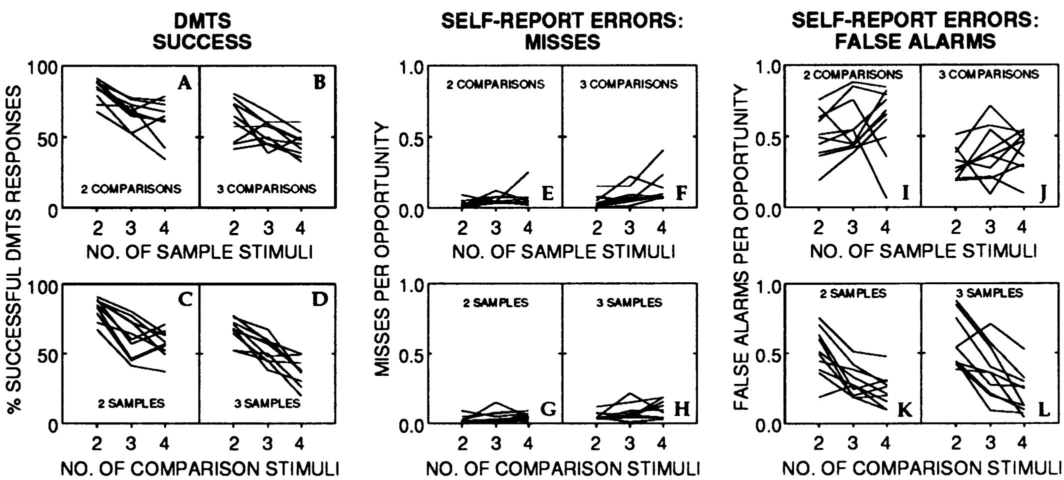

Fig. 6. DMTS success and rates of self-report errors as a function of the number of sample stimuli (top rows of panels) and the number of comparison stimuli (bottom rows of panels) in the DMTS task. Each panel contains one function per subject.

and D (those showing the influence of comparison-stimulus number). Although statistical comparison across manipulations is precluded by the fact that some experimental conditions appear in both analyses, visual comparison suggests that the functions in Panels A and B are roughly similar to those in Panels C and D, respectively. That is, manipulating the number of stimuli appeared to produce the same effect on overall DMTS success regardless of whether the manipulation took place at the level of sample or comparison stimuli. As a result, any differences in self-reports resulting from the manipulation of the number of sample and comparison stimuli cannot be attributed to gross differences in the way the two manipulations affected the referent performance.

The remaining portions of Figure 6 show rates of the two types of self-report errors—misses and false alarms—as a function of the number of DMTS sample and comparison stimuli. The middle panels show that miss rates tended to increase slightly in some subjects with the number of DMTS stimuli; this effect did not depend on whether sample or comparison stimuli were manipulated. The right panels show false-alarm rates. When the number of DMTS sample stimuli was manipulated (Panels I and J), patterns of false-alarm rates varied substantially from subject to subject. Some subjects showed increases in false-alarm rates between the one-sample and the three-sample condition, as found previously by Critchfield and Perone (1993), but there was no statistically significant group effect (Appendix). By contrast, when the number of comparison stimuli was manipulated (Panels K and L), a consistent pattern emerged, with false-alarm rates negatively correlated with the number of comparison stimuli.

Figure 7 (top panels) shows self-report bias as a function of the number of DMTS stimuli. As expected from the data in Figure 5, the bias scores tended to cluster near the negative end of the scale (the report-success bias described earlier). Bias also tended to become less pronounced as DMTS stimuli (samples or comparisons) became more numerous. The bottom panels in Figure 7 show self-report discriminability as a function of the number of DMTS sample and comparison stimuli. When the number of DMTS samples was manipulated (Panels E and F), discriminability tended to decrease; however, when the number of DMTS comparisons was manipulated (Panels G and H), discriminability tended to increase.

Relative Influence on Self-Reports of DMTS Speed and Accuracy

To summarize the results thus far, the sample-stimulus and comparison-stimulus manipulations appeared to have similar overall effects on DMTS success and on self-report bias,

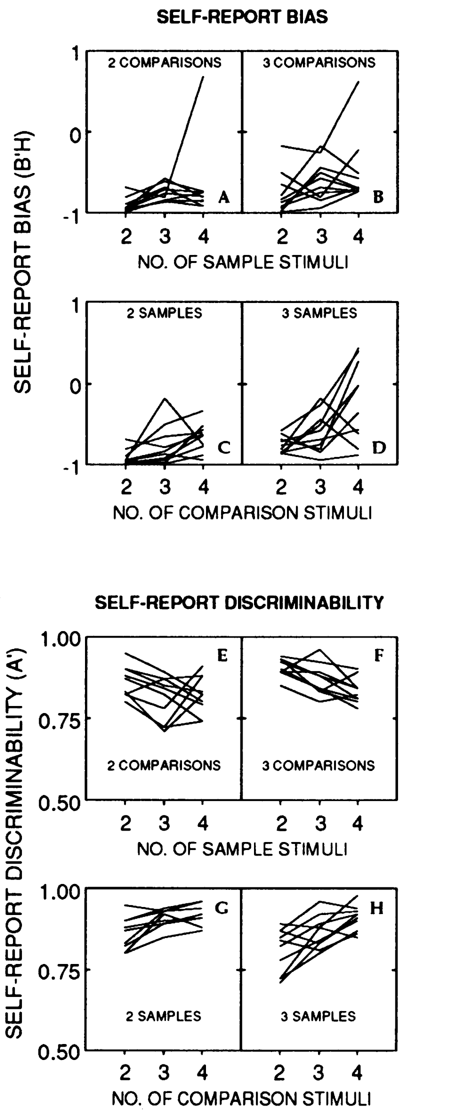

Fig. 7. Self-report bias and self-report discriminability as a function of the number of sample stimuli (top rows of panels) and the number of comparison stimuli (bottom rows of panels). Each panel contains one function per subject. For bias, positive scores indicate a predisposition for reporting failure; negative scores indicate a predisposition for reporting success.

but different effects on self-report discriminability. The remaining analyses consider these effects in the context of a more detailed examination of the DMTS performance about which subjects made their self-reports.

Figure 8 shows group-mean DMTS performance in terms of the two response characteristics—accuracy and speed—on which points were contingent (corresponding individual-subject data are shown in Table 3). The bottom left pair of panels shows mean rates of accurate DMTS matching, irrespective of whether the choices met the time limit. Horizontal lines indicate chance levels of accurate matching, defined by the number of comparison stimuli. Raw percentages of matching decreased similarly as the numbers of sample and comparison stimuli were increased; however, matching decreased more, relative to chance, under the sample-stimulus manipulation. To clarify this relationship, the bottom right pair of panels of Figure 8 shows rates of accurate matching adjusted for chance. Prior to averaging, individual accuracy scores in each experimental condition were converted to a percentage of possible improvement above chance, according to the following formula: 100 × [(% correct matches − chance) ÷ (100 − chance)]. The figure shows that the sample-stimulus manipulation affected adjusted accuracy scores more than the comparison-stimulus manipulation did. These patterns can be contrasted with those for DMTS speed, considered irrespective of matching accuracy (top right pair of panels of Figure 8). DMTS speed was influenced more by the comparison-stimulus manipulation than by the sample-stimulus manipulation. In sum, the sample-stimulus and comparison-stimulus manipulations tended to have different effects on the trade-off between speed and accuracy in DMTS performance.

Because the sample-stimulus and comparison-stimulus manipulations produced different patterns of DMTS speed and accuracy, Figure 9 examines whether self-reports were differentially sensitive to the speed and accuracy of DMTS choices. The influence of DMTS accuracy on self-reports ("accuracy-detection" bias and discriminability; filled triangles) was estimated by analyzing only those trials on which the DMTS response was faster than the time limit (on these trials, DMTS responses varied only in terms of their accuracy). The influence of DMTS speed on self-reports ("speed-detection" bias and discriminability; unfilled triangles) was estimated by analyzing only those trials on which the DMTS choice was accurate (on these trials, DMTS responses varied only in terms of their speed). These analyses indicate the relative influence of DMTS speed and accuracy on bias and discriminability (Critchfield & Perone, 1993), although the resulting absolute values of A′

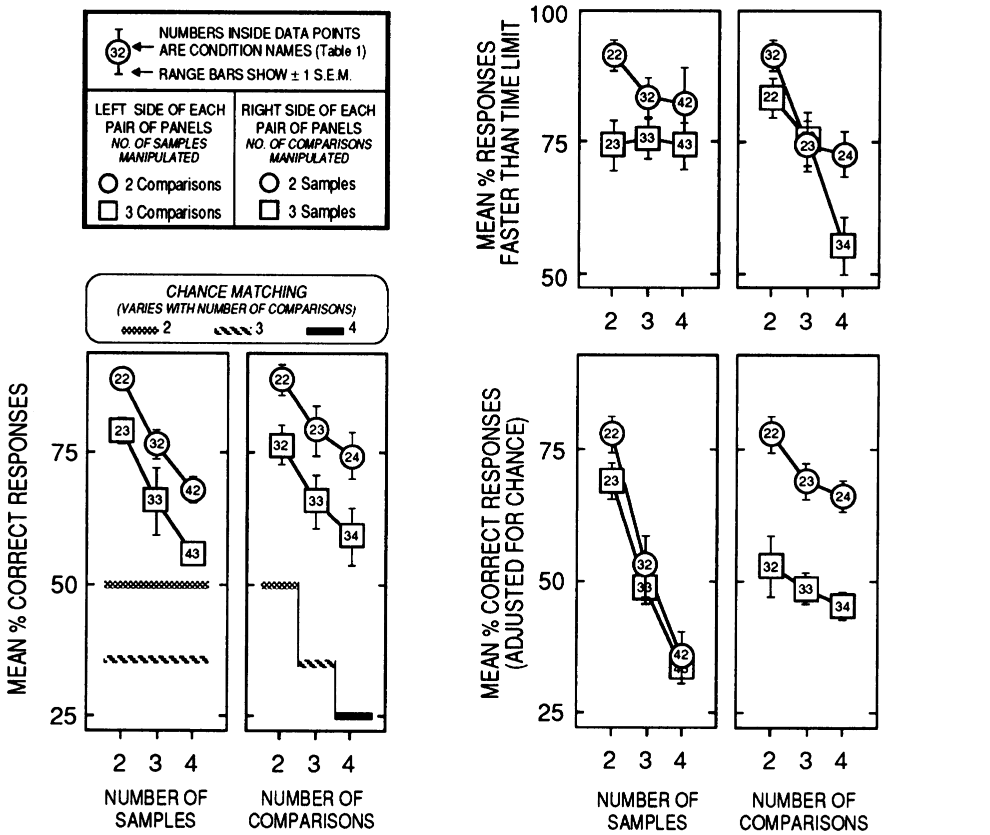

Fig. 8. Group-mean percentage of DMTS responses that met the accuracy or speed components of the point-reinforcement contingency. Bottom panels show the percentage of trials with an accurate match irrespective of speed. See text for description of calculations adjusting for chance accuracy. Top right pair of panels shows the percentage of trials with a response faster than the time limit, irrespective of accuracy. Range bars show ±1 *SEM*, except where data points exceed the size of the bars.

and B'_H are not especially informative due to the post hoc nature of the analysis and the fact that subjects made self-reports about their overall DMTS success rather than separate self-reports about DMTS speed and accuracy.

For simplicity of presentation, data from sample-stimulus manipulations were collapsed across levels of comparison stimuli, and data from comparison-stimulus manipulations were collapsed across levels of sample stimuli. Accuracy-detection bias tended to be more extreme than speed-detection bias. Analyses of variance in a two (speed vs. accuracy) by three (levels of stimuli) design revealed a statistically significant difference between speed-detection and accuracy-detection biases for the comparison-stimulus manipulation, $F(1, 9) = 5.7$, $p = .04$. A visually similar pattern was not statistically significant for the sample-stimulus manipulation, $F(1, 9) = 3.1$, $p = .11$.

Speed-detection discriminability tended to be higher than accuracy-detection discriminability. Analyses of variance revealed a significant difference between speed-detection and accuracy-detection discriminability for both the comparison-stimulus manipulations $F(1, 9) = 11.7$, $p = .008$, and sample-stimulus manipulations, $F(1, 9) = 27.1$, $p = .0006$. Speed-detection discriminability was not systematically related to number of sample stimuli, but was positively related to the number of comparison stimuli. Accuracy-detection discriminability was negatively related to the number of samples, but was positively related to the number of comparisons.

Figure 10 integrates the differential effects of sample-stimulus and comparison-stimulus manipulations on DMTS speed and accuracy with the differential sensitivity of self-reports to those response characteristics. Bias and dis-

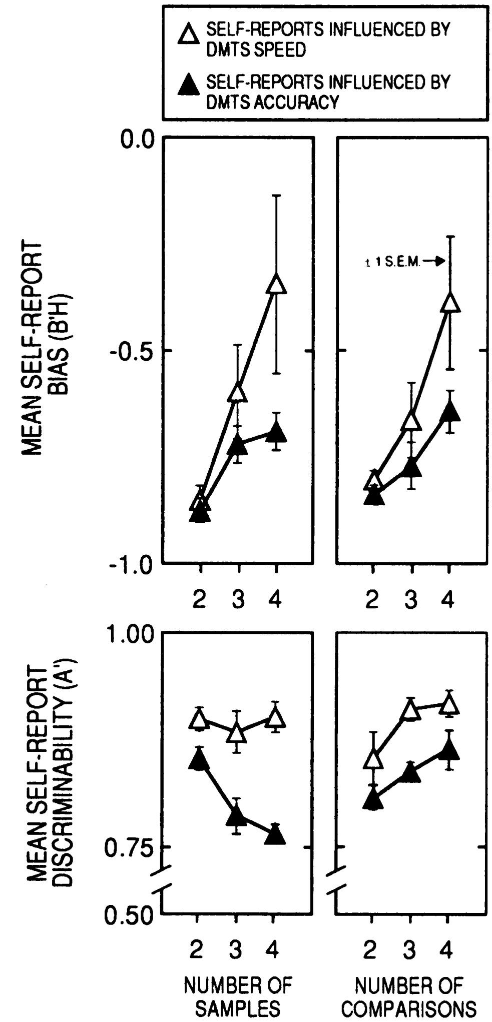

Fig. 9. Group-mean self-report bias and self-report discriminability as influenced by the speed and accuracy of DMTS referent responses. Range bars show ±1 *SEM*, except where data points exceed the size of the bars. See text for description of the calculation of bias and discriminability scores.

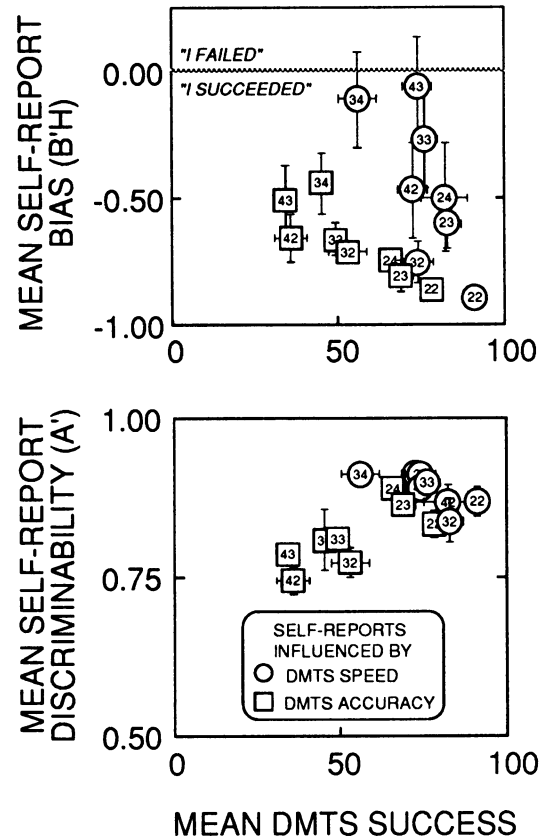

Fig. 10. Group-mean self-report bias and self-report discriminability, as influenced by the speed and accuracy of DMTS responses, plotted as a function of mean DMTS success. For self-reports influenced by response speed, the abscissa shows the percentage of trials with referent responses faster than the time limit. For self-reports influenced by response accuracy, the abscissa shows the adjusted percentage of trials with correct referent responses. See text for description of calculations adjusting for chance accuracy. Range bars show ±1 *SEM*, except where data points exceed the size of the bars.

criminability are plotted against rates of DMTS success. For self-reports influenced by DMTS speed (circles), the abscissa shows the percentage of trials with DMTS responses faster than the time limit. For self-reports influenced by DMTS accuracy (squares), the abscissa shows the percentage of trials with correct DMTS choices, adjusted for chance.

The top panel of Figure 10 shows that both speed-detection bias and accuracy-detection bias tended to become more extreme as DMTS success became more frequent. The slope for speed-detection bias was steeper than that for accuracy-detection bias, indicating more pronounced signal-frequency effects when speed was the referent-response characteristic at issue. (Because of the imprecise nature of the speed-detection and accuracy-detection estimates, the functions in Figure 10 can shed no light on the resistance to change in overall self-report bias identified in Figure 5.)

The bottom panel of Figure 10 shows discriminability effects consistent with those in previous figures. Condition names appear inside the data points. Accuracy-detection discriminability tended to increase with the number of comparison stimuli and decrease with the number of comparison stimuli. Speed-de-

Table 4

Summary of findings: Changes in self-reports coincident with increasing the number of DMTS distractor stimuli. Footnotes indicate whether relevant findings replicate those of Critchfield and Perone (1993). Effects not footnoted are new findings.

Dependent measure	Changes with increase in number of: Sample stimuli	Comparison stimuli	Other effects
DMTS errors	increase[a]	increase	
self-report errors			
misses	increase[a]	increase	
false alarms	no change[b]	decrease	
self-report bias	decrease[a]	decrease	ACC > SPEED[c] increase w/DMTS success[a]
self-report discriminability	decrease[a] SPEED: no change[b] ACC: decrease[a]	increase SPEED: increase ACC: increase	SPEED > ACC[d]

Note. ACC = self-reports influenced by DMTS accuracy; SPEED = self-reports influenced by DMTS speed. See text for description of the analyses.

[a] Replicates the previous study.

[b] Results of the previous study were variable, making comparison across studies difficult.

[c] The previous study found no differences between speed-detection bias and accuracy-detection bias.

[d] The previous study found accuracy-detection discriminability to be higher than speed-detection discriminability.

tection discriminability tended to increase with the number of comparison stimuli but not with the number of sample stimuli.

DISCUSSION

The results of this experiment both replicate and extend those of previous studies of verbal self-reports about human DMTS performance. The experiment identified five separate influences on self-reports (the number of DMTS sample and comparison stimuli, DMTS success rates, and the speed and accuracy of DMTS responses) and in some cases characterized their interaction. Table 4 summarizes the results and compares them with those of the related study by Critchfield and Perone (1993). That major effects of the previous study's sample-stimulus manipulation were replicated lends confidence to these findings as well as to new findings in the present study. The only discrepancies with the previous study—relative levels of self-report bias and discriminability associated with DMTS speed and accuracy—constitute an intriguing subject for future research but do not bear on interpretation of the other results.

Self-Report Bias

Signal frequency is a determinant of bias in psychophysical studies (Gescheider, 1985), and simple verbal self-reports are structurally similar to psychophysical judgments. The present experiment was designed to examine self-reports occurring coincident with a broad range of DMTS success rates to determine the role of signal frequency in self-report bias. As in previous studies, subjects usually exhibited a report-success bias (Critchfield & Perone, 1993). Importantly, the bias was most pronounced when successful DMTS responses were most frequent and least pronounced in conditions with infrequent DMTS success—an outcome consistent with signal-frequency effects. This pattern held when the data were adjusted for chance matching in the DMTS referent task as well as when the influences of DMTS speed and accuracy on bias were considered separately (Figure 10).

Perhaps because of procedural differences, the relationship between bias and DMTS success rates was stronger in the present study than in that of Critchfield and Perone (1993). It may be relevant, for example, that in the previous study DMTS difficulty was manipulated across trials within every session, whereas in the present study difficulty was manipulated across experimental conditions. Another procedural variable worthy of investigation is the point contingency on self-reports. Because accurate "I succeeded" and "I failed" self-reports produced points with the same conditional probability (.2), the point contingency in the present study would not be expected to affect self-report bias (Gescheider, 1985). Nevertheless, reinforcement frequency was not manipulated in the present study and merits investigation as a source of bias; certainly differential reinforcement probabilities for "I succeeded" and "I failed" self-reports would be expected to contribute to bias (McCarthy & Davison, 1981).

Signal-frequency effects may have influenced the relationship between self-report bias and DMTS success rate, but the frequency of "I succeeded" self-reports did not simply match the frequency of successful responses. Simple probability matching (e.g., Craig, 1976) predicts a shift from report-success bias, across a zero-bias threshold and toward a report-failure bias, when DMTS success occurred on 50% of the trials. Some subjects never crossed the threshold, and in no case did the shift occur unless DMTS success rates were below 50% (Figure 5), suggesting that signal frequency was not the sole source of self-report bias. This resistance to change remains to be explained, and without additional research procedural artifact cannot be ruled out as a contributing influence. For example, subjects received extensive preliminary training (several hundred DMTS trials) prior to the introduction of the self-report portion of the procedure. Because the DMTS task initially was less demanding than in the main experiment, preliminary training could have established a lasting predisposition to perceive DMTS responses as successful.

One reason for skepticism about artifact as a substantial source of bias is the apparent similarity of bias effects in the present study to patterns that regularly occur in other contexts (J. Brown, 1986; Cameron & Evers, 1990; Dobson & Franche, 1989; Halbreich et al., 1989; T. Nelson, McIntyre, LaBrie, & Csiky, 1991; Weinstein, 1980). For example, a large majority of American adults rate their automobile driving abilities as above average (Svenson, 1981), and American employees typically overrate their job performance compared to more objective measures (Murphy & Cleveland, 1991).

Patterns analogous to the report-success bias seen here are common enough that some theorists view self-enhancement as a generic human trait (Bjorkland & Green, 1992; Furnham, 1986; Taylor & Brown, 1988) that presumably sets the boundaries within which situational influences on self-evaluation can operate. The resistance to change of self-report biases in the present study (Figure 5) apparently conforms to predictions derivable from this point of view. Yet proposals appealing to human nature remain controversial in part due to ambiguities about the ultimate origins of the bias. Bjorkland and Green (1992), for example, argue for phylogenic origins based on hints of self-enhancement in the verbal behavior of very young children (e.g., Stipek, 1984; Yussen & Levy, 1975). By contrast, other theorists assume social origins of self-knowledge and self-descriptive verbal behavior (e.g., Carver & Scheier, 1981; Markus & Kitayama, 1991; Skinner, 1953, 1957; Stipek, Recchia, & McClintic, 1992; Vygotsky, 1978).

Substantial anecdotal evidence supports a social-acquisition model. For example, parents, as the primary verbal community early in life, might be one important social influence on self-evaluation. It is interesting to note, therefore, that American parents consistently overestimate their children's abilities (Gretarsson & Gelfand, 1988; Heriot & Schmickel, 1967; Miller, Manhal, & Mee, 1991). Moreover, if self-description is socially acquired, then it also should be culturally variable, given that cultural practices comprise a substantial portion of the local social environment. Compared to Americans, Poles and Swedes tend to rate their driving skill more modestly (Goszczyńska & Rosłan, 1989; Svenson, 1981), and Chinese employees tend to rate their job performance more realistically (Fahr, Dobbins, & Cheng, 1991). Similarly, American students, who score relatively low on mathematics achievement tests, tend to judge their math skills positively, whereas Chinese students, who score relatively high, tend to judge their skills less positively (Stevenson, Chen, & Lee, 1993; Stevenson, Lee, & Stigler, 1986). Consistent with a social-acquisition model, parents' descriptions of child math achievement in China and the United States generally correspond to those of their children (Chen & Uttal, 1988; Stevenson et al., 1990). It is important to note, however, that cultural differences in self-evaluation have been measured only broadly with survey techniques and only with respect to a limited number of referents, such as work and academic performance—self-descriptions of which might be idiosyncratically linked to variations in cultural verbal practices. Procedures like those of the present study, which manipulate referent performances chosen for experimental utility rather than social importance, would provide a useful test of the generality of cultural differences in self-evaluation.

Appeals to social and cultural influences, although consistent with an ontogenic perspective on self-evaluation, ultimately beg the critical question of what behavioral mechanisms are responsible for the development of self-report biases in individuals, because for social and cultural practices to be transmitted there must first exist individuals who engage in them. Existing explanations typically invoke constructs such as self-efficacy (Bandura, 1989), defense mechanisms (Sackeim, 1983), and egocentrism (Bjorkland & Green, 1992), although Skinner (1953, 1957) proposed two possible mechanisms, based on the three-term operant contingency, that may provide direction for future research. First, self-reports may become biased in a manner consistent with the present results when the social contingencies normally governing them are supplemented by "automatic" reinforcing and punishing effects (Skinner, 1957, pp. 163–166 and chap. 15; see also Vaughan & Michael, 1982). Second, stimulus control over self-reports can be selectively disrupted by intermittent social reinforcement of self-reports that do not correspond to their presumed referents (Skinner, 1953, pp. 258–261, 1957, pp. 130–135 and 147–151). Signal-detection studies offer precedent for the latter suggestion (McCarthy & Davison, 1981), and both explanations are appealing in their consistency with well-established behavioral principles, but neither has been systematically investigated.

Self-Report Discriminability

In a previous study of self-reports about matching-to-sample success, Critchfield and Perone (1993) found that self-report discriminability decreased as DMTS success rates decreased. The present study isolated the source of this correlation in the number and type of stimuli in the DMTS referent task. When DMTS success was manipulated via the number of sample stimuli on each trial, as in the previous study, self-report discriminability decreased. When DMTS success was decreased via the number of comparison stimuli, however, self-report discriminability increased.

The sample-stimulus and comparison-stimulus manipulations thus appeared to have opposite effects on self-report discriminability, but the two manipulations also produced qualitatively different patterns of referent-task performance. Superficially similar changes in DMTS success rates (Figure 6) proved to derive primarily from accuracy decrements in the case of the sample-stimulus manipulation and from speed decrements in the case of the comparison-stimulus manipulation (Figure 8). To determine whether self-report discriminability was an artifact of these patterns, the data were reanalyzed to consider the role of chance rates of DMTS matching and the separate influences of DMTS speed and accuracy on self-reports (Figures 9 and 10). Differential effects remained intact, indicating that the two manipulations indeed had different effects on self-report discriminability separate from their effects on DMTS performance.

The reasons for differential discriminability effects remain unclear, although the DMTS procedure used to generate referent responses will offer one avenue of inquiry for future studies. For example, ITI duration may be of interest because of its role in reducing "interference" by events from previous trials (Mackay, 1991; Maki, Moe, & Bierley, 1977; Wright, Urcuioli, & Sands, 1986). The present experiment used an ITI lasting only 1 s, placing sample stimuli relatively close in time to the previous trial (samples necessarily occur closer to the previous trial than do comparisons), where interference effects could interact with the competing stimulus control created by increasing the number of sample stimuli. It is possible, therefore, that increasing the duration of the ITI would weaken or eliminate the negative relationship between self-report discriminability and the number of DMTS sample stimuli. Also worth investigating is the duration of the retention interval, which separates sample stimuli (but not comparison stimuli) in time from the self-report.

Conclusions

Bias and discriminability indices adopted from signal-detection theory can serve a valuable role in the experimental analysis of verbal self-reports. Their use in the present study showed self-report bias and discriminability to be influenced by different characteristics of a target DMTS response. Bias appeared to reflect the interaction of two factors, a situational influence analogous to signal-frequency effects in psychophysical studies and possibly an extra experimental influence whose parameters and sources remain to be explored. By contrast, self-report discriminability was more clearly related to two aspects of the DMTS task that produced referent responses (i.e., the sample- and comparison-stimulus configurations).

That neither bias nor discriminability effects could be inferred from gross accuracy scores (e.g., Figure 3) highlights the utility of signal-detection indices in suggesting sources of control over verbal self-reports. Analogous measures have aided the analysis of self-reports in a variety of research traditions. Examples include reports by clients with eating disorders of their body shape and size (Gardner, Martinez, & Espinoza, 1987), reports by students on the accuracy of their exam answers (Hosseini & Ferrell, 1982), and reports by subjects of "private" events such as heart beats (Pennebaker & Hoover, 1984), minuscule muscle twitches (Hefferline & Perera, 1963), and drug sensations (Colpaert, 1978). As a cautionary note, strict application of signal-detection theory requires assumptions (e.g., stability of decision criteria) that may not be met in studies like the present one (Colpaert, 1987; Green & Swets, 1966). Nevertheless, there is ample precedent for applying analytical strategies of signal-detection theory without necessarily adopting the theoretical framework on which it was originally based (e.g., McCarthy & Davison, 1981; Nevin, 1981).

Although verbal behavior may be difficult to analyze in its natural state, laboratory procedures can facilitate the study of phenomena (e.g., self-enhancement) and controlling variables (e.g., signal frequency) that are of potential importance in self-reports outside the laboratory. The procedures of the present study originally were devised to permit the study of self-reports under relatively simple conditions (Critchfield & Perone, 1993), but, even so, many variables still require empirical attention. These range from procedural factors (e.g., between- vs. within-session manipulation of trial difficulty) to the immediate consequences of self-reports (both programmed and "automatic") to historical variables (e.g., cultural influences). The kind of systematic experimentation necessary to examine this range of variables often reveals unexpected complexity in behavior (e.g., Perone & Courtney, 1992; Sidman & Tailby, 1982), and if the present results are any indication, there is likely to be nothing simple about "simple" self-reports.

REFERENCES

Bandura, A. (1989). Regulation of cognitive processes through perceived self-efficacy. *Developmental Psychology,* **25,** 729–735.

Baron, A., & Menich, S. R. (1985). Reaction times of older and younger men: Effects of compound samples and a prechoice signal on delayed matching-to-sample performances. *Journal of the Experimental Analysis of Behavior,* **44,** 1–14.

Bjorkland, D. F., & Green, B. L. (1992). The adaptive nature of cognitive immaturity. *American Psychologist,* **47,** 46–54.

Brown, J. D. (1986). Evaluations of self and others: Self-enhancement biases in social judgments. *Social Cognition,* **4,** 353–376.

Brown, R. (1970). *Psycholinguistics: Selected papers.* New York: Free Press.

Cameron, R., & Evers, S. (1990). Self-report issues in obesity and weight management: State of the art and future directions. *Behavioral Assessment,* **12,** 91–106.

Carver, C. S., & Scheier, M. F. (1981). *Attention and self-regulation: A control-theory approach to human behavior.* New York: Springer-Verlag.

Chen, C., & Uttal, D. A. (1988). Cultural values, parents' beliefs, and children's achievement in the United States and China. *Human Development,* **31,** 351–358.

Colpaert, F. C. (1978). Some properties of drugs as physiological signals: The FR procedures and signal detection theory. In F. C. Colpaert & J. L. Slangen (Eds.), *Stimulus properties of drugs: Ten years of progress* (pp. 217–242). Amsterdam: Elsevier/North Holland.

Colpaert, F. C. (1987). Drug discrimination: Methods of manipulation, measurement, and analysis. In M. A. Bozarth (Ed.), *Methods of assessing the reinforcing properties of abused drugs* (pp. 341–377). New York: Springer-Verlag.

Craig, A. (1976). Signal recognition and the probability-matching decision rule. *Perception & Psychophysics,* **20,** 157–162.

Critchfield, T. S., & Perone, M. (1990a). Verbal self-reports as a function of speed, accuracy and reinforcement of the reported performance. *Psychological Record,* **40,** 541–554.

Critchfield, T. S., & Perone, M. (1990b). Verbal self-reports of delayed matching to sample by humans. *Journal of the Experimental Analysis of Behavior,* **53,** 321–344.

Critchfield, T. S., & Perone, M. (1993). Verbal self-reports about matching to sample: Effects of the number of elements in a compound sample stimulus. *Journal of the Experimental Analysis of Behavior,* **59,** 193–214.

Dobson, K., & Franche, R. (1989). A conceptual and empirical review of the depressive realism hypothesis. *Canadian Journal of Behavioral Science,* **21,** 419–433.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data.* Cambridge, MA: MIT Press.

Fahr, J., Dobbins, G. H., & Cheng, B. (1991). Cultural relativity in action: A comparison of self-ratings made by Chinese and U.S. workers. *Personnel Psychology,* **44,** 129–147.

Furnham, A. (1986). Response bias, social desirability, and dissimulation. *Personality and Individual Differences,* **7,** 385–400.

Gardner, R. M., Martinez, R., & Espinoza, T. (1987). Psychophysical measurement of body image of self and others in obese subjects. *Journal of Social Behavior and Personality,* **2,** 205–217.

Geisser, S., & Greenhouse, S. W. (1958). An extension of Box's results on the use of the F distribution in multivariate analysis. *Annals of Mathematical Statistics,* **29,** 149–158.

Gescheider, G. A (1985). *Psychophysics: Method, theory, and application.* Hillsdale, NJ: Erlbaum.

Goszczyńska, M., & Roslan, A. (1989). Self-evaluation of drivers' skill: A cultural comparison. *Accident Analysis and Prevention,* **21,** 217–224.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Gretarsson, S. J., & Gelfand, D. M. (1988). Mothers' attributions regarding their children's social behavior and personality characteristics. *Developmental Psychology,* **24,** 264–269.

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin,* **75,** 424–429.

Halbreich, U., Bakhai, Y., Bacon, K. B., Goldstein, S., Asnis, G. M., Endicott, J., & Lesser, J. (1989). The normalcy of self-proclaimed "normal volunteers." *American Journal of Psychiatry,* **146,** 1052–1055.

Hayes, S. C. (1986). The case of the silent dog—Verbal reports and the analysis of rules: A review of Ericsson and Simon's *Protocol Analysis: Verbal Reports as Data. Journal of the Experimental Analysis of Behavior,* **45,** 351–363.

Hefferline, R. F., & Perera, T. B. (1963). Proprioceptive discrimination of a covert operant without its observation by the subject. *Science,* **139,** 834–835.

Heriot, J. T., & Schmickel, G. A. (1967). Maternal estimates of IQ in children evaluated for learning potential. *American Journal of Mental Deficiency,* **71,** 920–924.

Hosseini, J., & Ferrell, W. R. (1982). Detectability of correctness: A measure of knowing that one knows. *Instructional Science,* **11,** 113–127.

Kausler, D. H., & Phillips, P. L. (1988). Instructional variation and adult age differences in activity memory. *Experimental Aging Research,* **14,** 195–199.

Klima, E. S., & Bellugi, U. (1979). *The signs of language.* Cambridge, MA: Harvard University Press.

Lennenberg, E. H. (1967). *Biological foundations of language.* New York: Wiley.

Mackay, H. A. (1991). Conditional stimulus control. In I. H. Iversen & K. A. Lattal (Eds.), *Experimental analysis of behavior* (Part 1, pp. 301-350). New York: Elsevier.

Maki, W. S., Moe, J. C., & Bierley, C. M. (1977). Short-term memory for stimuli, responses, and reinforcers. *Journal of Experimental Psychology: Animal Behavior Processes,* **3,** 156–177.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review,* **98,** 224–253.

McCarthy, D., & Davison, M. (1981). Matching and signal detection. In M. L. Commons & J. A. Nevin (Eds.), *Quantitative analysis of behavior: Vol. 1. Discriminative properties of reinforcement schedules* (pp. 393–417). Cambridge, MA: Ballinger.

McPherson, A., Bonem, M., Green, G., & Osborne, J. G. (1984). A citation analysis of the influence on research of Skinner's *Verbal Behavior. Behavior Analyst,* **7,** 157–167.

Miller, S. A., Manhal, M., & Mee, L. L. (1991). Parental beliefs, parental accuracy, and children's cognitive performance: A search for causal relations. *Developmental Psychology,* **27,** 267–276.

Murphy, K. R., & Cleveland, J. N. (1991). *Performance appraisal: An organizational perspective.* Boston: Allyn and Bacon.

Nelson, R. O. (1977). Assessment and therapeutic functions of self-monitoring. In M. Hersen, R. M. Eisler, & P. M. Miller (Eds.), *Progress in behavior modification* (Vol. 5, pp. 263-308). New York: Academic Press.

Nelson, T. M., McIntyre, J. W. R., LaBrie, I. G., & Csiky, A. (1991). Self-perception of ability to work in the cold. *Behavioral Medicine,* **17,** 15–23.

Nevin, J. A. (1981). Psychophysics and reinforcement schedules: An integration. In M. L. Commons & J. A. Nevin (Eds.), *Quantitative analysis of behavior: Vol. 1. Discriminative properties of reinforcement schedules* (pp. 3–30). Cambridge, MA: Ballinger.

Oah, S., & Dickinson, A. M. (1989). A review of empirical studies of verbal behavior. *Analysis of Verbal Behavior,* **7,** 53–68.

Pennebaker, J. W., & Hoover, C. W. (1984). Visceral perception versus visceral detection: Disentangling methods and assumptions. *Biofeedback and Self-Regulation,* **9,** 339–352.

Perone, M., & Courtney, K. (1992). Fixed-ratio pausing: Joint effects of past reinforcer magnitude and stimuli correlated with upcoming magnitude. *Journal of the Experimental Analysis of Behavior,* **57,** 33–46.

Sackeim, H. A. (1983). Self-deception, self-esteem, and depression: The adaptive value of lying to oneself. In J. Masling (Ed.), *Empirical studies of psychoanalytic theories* (Vol. 1, pp. 101–157). Hillsdale, NJ: Analytic Press.

Shimp, C. P. (1981). The local organization of behavior: Discrimination of and memory for simple behavioral patterns. *Journal of the Experimental Analysis of Behavior,* **36,** 303–315.

Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior,* **37,** 5–22.

Skinner, B. F. (1953). *Science and human behavior.* New York: Free Press.

Skinner, B. F. (1957). *Verbal behavior.* New York: Appleton-Century-Crofts.

Stevenson, H. W., Chen, C., & Lee, S. (1993). Mathematics achievement of Chinese, Japanese, and American children: Ten years later. *Science,* **259,** 53–58.

Stevenson, H. W., Lee, S., Chen, C., Lummis, M., Stigler, J., Fan, L., & Ge, F. (1990). Mathematics achievement of children in China and the United States. *Child Development,* **61,** 1053–1066.

Stevenson, H. W., Lee, S., & Stigler, J. W. (1986). Mathematics achievement of Chinese, Japanese, and American children. *Science,* **231,** 693–699.

Stipek, D. (1984). Young children's performance expectations: Logical analysis or wishful thinking? In J. G. Nicholls (Ed.), *Advances in motivation and achieve-*

ment: Vol. 3. The development of achievement motivation (pp. 33-56). Greenwich, CT: JAI Press.

Stipek, D., Recchia, S., & McClintic, S. (1992). Self-evaluation in young children. *Monographs of the Society for Research in Child Development*, **57** (1, Serial No. 226).

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica*, **47**, 143-148.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, **103**, 193-210.

Vaughan, M. E., & Michael, J. L. (1982). Automatic reinforcement: An important but ignored concept. *Behaviorism*, **10**, 217-227.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds. & Trans.). Cambridge, MA: Harvard University Press.

Walker, C. J., & Blaine, B. (1991). The virulence of dread rumors: A field experiment. *Language & Communication*, **11**, 291-297.

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, **39**, 806-820.

Wright, A. A., Urcuioli, P. J., & Sands, S. F. (1986). Proactive interference in animal memory. In D. F. Kendrick, M. E. Rilling, & M. R. Denny (Eds.), *Theories of animal memory* (pp. 101-125). Hillsdale, NJ: Erlbaum.

Yussen, S., & Levy, V. (1975). Developmental changes in predicting one's own span of memory. *Journal of Experimental Child Psychology*, **19**, 502-508.

Received October 6, 1992
Final acceptance July 6, 1993

APPENDIX

Results of repeated measures analyses of variance, with number of DMTS sample stimuli and number of DMTS comparison stimuli as within-subject factors, conducted to corroborate the graphic displays in Figures 6 and 7. Each effect was tested in two separate 3 by 2 designs derived from the list of experimental conditions in Table 1; see text for details. Analyses employ the Geisser and Greenhouse (1958) correction.

Dependent variable and type of effect	Repeated measures ANOVA model	
	3 samples by 2 comparisons	3 comparisons by 2 samples
DMTS success		
Main: samples	$F(2, 9) = 41.2, p = .0001$	$F(1, 9) = 40.0, p = .0001$
Main: comparisons	$F(1, 9) = 73.9, p = .0001$	$F(2, 9) = 95.6, p = .0001$
Interaction	$F(2, 9) = 1.2, p = .321$	$F(2, 9) = 2.4, p = .117$
Self-report errors: misses		
Main: samples	$F(2, 9) = 15.8, p = .0003$	$F(1, 9) = 6.9, p = .0001$
Main: comparisons	$F(1, 9) = 5.6, p = .042$	$F(2, 9) = 48.5, p = .011$
Interaction	$F(2, 9) = 0.4, p = .386$	$F(2, 9) = 0.5, p = .555$
Self-report errors: false alarms		
Main: samples	$F(2, 9) = 2.2, p = .144$	$F(1, 9) = 1.0, p = .337$
Main: comparisons	$F(1, 9) = 24.6, p = .0008$	$F(2, 9) = 32.7, p = .0001$
Interaction	$F(2, 9) = 0.0, p = .939$	$F(2, 9) = 3.2, p = .068$
Self-report bias		
Main: samples	$F(2, 9) = 9.0, p = .006$	$F(1, 9) = 11.4, p = .008$
Main: comparisons	$F(1, 9) = 3.2, p = .110$	$F(2, 9) = 18.0, p = .0006$
Interaction	$F(2, 9) = 0.0, p = .995$	$F(2, 9) = 1.8, p = .195$
Self-report discriminability		
Main: samples	$F(2, 9) = 10.8, p = .002$	$F(1, 9) = 10.1, p = .011$
Main: comparisons	$F(1, 9) = 20.6, p = .001$	$F(2, 9) = 22.0, p = .0002$
Interaction	$F(2, 9) = 0.6, p = .472$	$F(2, 9) = 4.6, p = .025$